International Journal of Production Research
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tprs20

Measuring production yield for processes with multiple quality characteristics

W. L. Pearn [a] , F. K. Wang [b] & C. H. Yen [a]
[a] Department of Industrial Engineering and Management , National Chiao Tung University , Taiwan, ROC
[b] Department of Industrial Management , National Taiwan University of Science and Technology , Taiwan, ROC
Published online: 22 Feb 2007.

International Journal of Production Research,
Vol. 44, No. 21, 1 November 2006, 4649–4661

Taylor & Francis
Taylor & Francis Group

Measuring production yield for processes with multiple quality characteristics

W. L. PEARN†, F. K. WANG*‡ and C. H. YEN†

†Department of Industrial Engineering and Management,
National Chiao Tung University, Taiwan, ROC
‡Department of Industrial Management,
National Taiwan University of Science and Technology, Taiwan, ROC

(*Revision received January 2006*)

Process yield is an important criterion used in the manufacturing industry for measuring process performance. Methods for measuring yield for processes with single characteristic have been investigated extensively. However, methods for measuring yield for processes with multiple characteristics have been comparatively neglected. In this paper, we develop a generalized yield index, called $TS_{pk,PC}$, based on the index S_{pk} introduced by Boyles (*Journal of Quality Technology*, **23**, 17–26, 1991) using the principal component analysis (PCA) technique. We obtained a lower confidence bound (LCB) for the true process yield. The proposed method can be used to determine whether a process meets the preset yield requirement, and make reliable decisions. Examples are provided to demonstrate the proposed methodology.

Keywords: Process yield; Process capability indices; Lower confidence bound; Principal component analysis

1. Introduction

Process capability indices, which establish the relationship between the actual process performance and the manufacturing specifications, have been the focus in quality assurance and capability analysis for the past 15 years. Those capability indices quantifying process performance are essential to any successful quality improvement activities and quality program implementation. The capability indices, C_p, C_{pk} and C_{pm}, are widely used in the manufacturing industry to evaluate process performance for cases with a single quality characteristic. The index C_p measures the overall process variation relative to the specification tolerance. The index C_{pk} takes into account the magnitude of process variation as well as the degree of process centering. The index C_{pm} emphasizes measuring the ability of process to cluster around the target, which reflects the degrees of process targeting. On the other hand, the index S_{pk} (Boyles 1991) is introduced to establish the relationship between the manufacturing specification and the actual process performance, which provides an exact measure on the process yield. Capability calculations for processes with single

*Corresponding author. Email: fukwun@mail.ntust.edu.tw

International Journal of Production Research
ISSN 0020–7543 print/ISSN 1366–588X online
http://www.tandf.co.uk/journals
DOI: 10.1080/00207540600589119

Table 1. The summary of multivariate capability indices.

Authors	Index	Distribution application	Tolerance form
Chan *et al.* (1991)	C_{pm}	Multivariate normal	Elliptical
Pearn *et al.* (1992)	${}_vC_{pm}$, ${}_vC_p$	Multivariate normal	Elliptical
Taam *et al.* (1993)	MC_{pm}, MC_p	Multivariate normal	Elliptical
Chen (1994)	MC_p	No specific	No specific
Shahriari *et al.* (1995)	$[C_{pM}, PV, LI]$	Multivariate normal	Elliptical
Wang and Du (2000)	MC_{pm}, MC_{pk}, MC_p	No specific	No specific

characteristic have been investigated extensively. Kotz and Johnson (2002) presented a thorough review for the development of process capability indices from 1992 to 2000.

Often, a manufactured product is described in multiple characteristics. That is, manufactured items require values of several different characteristics for adequate description of their quality. Each of those characteristics must satisfy certain specifications. The assessed quality of a product depends on the combined effects of those characteristics, rather than on their individual values. For example, automobile paint needs a range of light reflective abilities and a range of adhesion abilities. A paint that satisfies one criterion but not the other is undesirable. Those characteristics are related through the compositions of the paint. It is therefore natural to consider a bivariate characterization of this paint. As for the tolerance region of multiple characteristics, we often take an ellipsoidal region or a rectangular region. In the two-dimension cases, those tolerance ranges compose a rectangular tolerance region. In higher dimensions, they form a hypercube. For more complex engineering specifications, the tolerance region is very complicated. For instance, a drawing of a connecting rod in a combustion engine consists of crank-bore inner diameter, pin-bore inner diameter, rod length, bore true-location and so on.

In order to handle the issue for cases with multiple quality characteristics, multivariate methods for assessing process capability are proposed. These relevant multivariate capability indices can be found in Chan *et al.* (1991); Pearn *et al.* (1992); Taam *et al.* (1993); Chen (1994); Shahriari *et al.* (1995); Wang and Du (2000), etc. A brief summary for multivariate capability indices is given in table 1. The multivariate capability indices proposed by Chan *et al.* (1991), Pearn *et al.* (1992), Taam *et al.* (1993) and Shahriari *et al.* (1995), respectively, require the assumption of multivariate normality while those proposed by Chen (1994) and Wang and Du (2000) make no assumption on multivariate normality. The tolerance regions of those methods using multivariate normality assumption are ellipsoidal except for Shahriari *et al.* (1995). Relatively, Chen (1994) and Wang and Du (2000) provide more flexible methods to assess the capability for multivariate data. Chen (1994) proposed this over a general tolerance zone which includes ellipsoidal and rectangular solid ones, and this manner does not rely on a particular distribution. Wang *et al.* (2000) presented a comparison of three methods proposed by Taam *et al.* (1993), Shahriari *et al.* (1995) and Chen (1994). Also, Wang and Du (2000) applied the principal component analysis (PCA) to process capability indices to handle normal and non-normal data. However, the issues between process yield and the multivariate capability indices have received little attention. In this paper, we focus on the process yield for the correlated multiple quality characteristics. We calculate the process yield using S_{pk} through PCA for processes with correlated multiple quality characteristics. We present the PCA method and the procedure of obtaining the lower confidence bound (LCB) for the true process yield using S_{pk} through the principal component analysis (PCA). Illustrative examples are given to demonstrate the applicability of the proposed approach.

2. Principal Component Analysis

PCA is a useful statistical technique that has been widely applied to face recognition and image compression, which is a common technique for finding patterns in high dimensional data. It is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. In many cases the patterns in data can be difficult to find in high-dimensional applications, particularly when graphical representation is not available, and PCA is a powerful tool in such situations. The other main advantage of PCA is that after finding patterns in the data, one could compress the data by reducing the number of dimensions without losing much information. PCA is a multivariate technique in which a number of related variables are transformed to a set of uncorrelated linear functions of the original measurements. The first principal component linearly combines all of the original variables in which the maximum variation among the objects is displayed. The second, third, and further components are, similarly, the linear combinations representing the next largest variation, irrespective of those represented by previous ones. In most practical applications, analysing the major components can retain most of the information regarding the variability of the process. In general, multivariate methods often assume the data satisfy multivariate normal distribution. But in applying the PCA technique one does not require such assumption.

Assume that X is a $\nu \times n$ sample data matrix, where ν is the number of product quality characteristic from one part and n is the sample size of part measured. Also, $\overline{X}$ is the sample mean vector ($\nu \times 1$) of observations and S is a $\nu \times \nu$ symmetric matrix representing the covariance between observations. Engineering specifications are given for each quality characteristic, where LSL and USL are their ν-vectors of the lower specification limits and upper specification limits, respectively. The vector T ($\nu \times 1$) represents the target values of the ν quality characteristics. In addition, the spectral decomposition can be used to obtain $D = U^T SU$, where D is a diagonal matrix. The diagonal elements of D, $\lambda_1, \lambda_2, \ldots, \lambda_\nu$, are the eigenvalues of S and the columns of U, $u_1, u_2, \ldots, u_\nu$ are the eigenvectors of S. Consequently, the *ith* principal component (PCi) is expressed as

$$PC_i = u_i^T x, \ \ \forall i = 1, 2, \ldots, \nu \tag{1}$$

where x is $\nu \times 1$ vectors on the original variables. The engineering specifications and target values of $PC_i s$ are as follows:

$$\begin{cases} LSL_{PC_i} = u_i^T LSL \\ USL_{PC_i} = u_i^T USL \quad \forall i = 1, 2, \ldots, \nu \\ T_{PC_i} = u_i^T T \end{cases} \tag{2}$$

Similarly, the relevant sample estimators, S^2 and $\overline{X}$ of $PC_i s$ can defined as

$$\begin{cases} S^2_{PC_i} = \lambda_i \\ \bar{X}_{PC_i} = u_i^T \bar{X} \end{cases} \quad \forall\, i = 1, 2, \ldots, v \tag{3}$$

The ratio of each eigenvalue to the summation of the eigenvalues is the proportion of variability associated with each principal component variable. That is,

$$\lambda_i \Big/ \sum_{i=1}^{v} \lambda_i, \quad \forall\, i = 1, 2, \ldots, v \tag{4}$$

However, only a few principal components can explain most of the total variability (about 80–90%). Anderson (1963) proposed a χ^2 test for identifying the significant components. It is

$$\chi^2 = -(n-1) \sum_{j=k+1}^{v} \ln \lambda_j + (n-1)(v-k) \ln \frac{\sum_{j=k+1}^{v} \lambda_j}{v-k} \tag{5}$$

where χ^2 has $r = (1/2)(v-k)(v-k+1) - 1$ degrees of freedom. Jackson (1980) further applied the test to the hypothesis H_0: $\lambda_{k+1} = \cdots = \lambda_v$ against the alternatives with at least one different eigenvalue. Referring to this method, we can choose the suitable number of $PC_i s$ rightly.

3. Process yield

Process yield has been the most basic and common criterion used in the manufacturing industry for measuring process performance. It is closely related to the production cost as well as customer satisfaction. Process yield is currently defined as the percentage of processed product unit passing inspection. That is, the product characteristic must fall within the manufacturing tolerance. For product units rejected (non-conformities), additional costs would be incurred to the factory for scrapping or repairing the product. All passed product units are equally accepted by the producer, which incurs the factory no additional cost. For processes with high yield, it produces few percentages of non-conforming products. That is, most of the products produced in this process satisfy the requirement of specifications. In many cases, a benchmark of minimum 99.73% for assessing the process is suggested. Enterprises get more profit and cost down with high process yield, hence companies make their efforts to increase the process yield. The relationships between the process yield and the process capability indices have been discussed extensively for processes with single characteristics, but comparatively neglected for processes with multiple characteristics.

Consider a production process in which, possibly dependent, quality characteristics determine the quality of the product. In other words, the product has multiple correlated characteristics. We are concerned with the probability of producing a good product satisfying all its specifications. Assume that the observations X have a multivariate normal distribution, $N_v(\mu, \sum)$, where v is the dimension of variables, μ is the mean vector and Σ represents the variance–covariance matrix of X. The components of the vectors LSL and USL are the v lower and upper specification limits, respectively. Under the assumptions mentioned, the probability that a production process produces a good product is

$$p = \int_{[LSL,USL]} N_v(X|\mu, \Sigma)dX \tag{6}$$

It is also called the true process yield.

For normally distributed processes, the index S_{pk} is used to establish the relationship between the manufacturing specification and the actual process performance, which provides an exact measure on the process yield, defined as

$$\begin{aligned} S_{pk} &= \frac{1}{3}\Phi^{-1}\left\{\frac{1}{2}\Phi\left(\frac{USL-\mu}{\sigma}\right)+\frac{1}{2}\Phi\left(\frac{\mu-LSL}{\sigma}\right)\right\} \\ &= \frac{1}{3}\Phi^{-1}\left\{\frac{1}{2}\Phi\left(\frac{1-C_{dr}}{C_{dp}}\right)+\frac{1}{2}\Phi\left(\frac{1+C_{dr}}{C_{dp}}\right)\right\} \end{aligned} \tag{7}$$

where $C_{dr} = (\mu - m)/d$, $C_{dp} = \sigma/d$, $m = (USL + LSL)/2$, $d = (USL - LSL)/2$. It provides an exact measure of process yield. If $S_{pk} = c$, then the process yield can be expressed as Yield $= 2\Phi(3c) - 1$. Obviously, there is a one-to-one correspondence between S_{pk} and the process yield. Considering processes with multiple characteristics, Chen *et al*. (2003) defined the yield index as

$$S_{pk}^T = \frac{1}{3}\Phi^{-1}\left\{\left[\prod_{j=1}^{v}\left(2\Phi(3S_{pkj})-1\right)+1\right]\Big/2\right\}, \tag{8}$$

where S_{pki} denote the S_{pk} value of the jth characteristic for $j = 1, 2, \ldots, v$ and v is the number of characteristics. The asymptotic distribution for an estimate $\hat{S}_{pk}^T$ can be found from Theorem 1 (see Appendix). This index provides an exact measure of the overall process yield when the characteristics are mutually independent. Also the overall process yield P can be established as

$$P = \prod_{j=1}^{v}\left[2\Phi(3S_{pkj})-1\right] = 2\Phi(3S_{pk}^T) - 1. \tag{9}$$

Assume that the multivariate processes data are from a multivariate normal distribution. In this case, the principal components can be applied to the capability study. Consequently, the new variables (principal components) are mutually independent and normal distributed (see Theorem 2 in the Appendix). Applying equation (8), the combined yield index for the multivariate processes data can be determined by

$$TS_{pk,PC} = \frac{1}{3}\Phi^{-1}\left\{\left[\prod_{j=1}^{v}\left(2\Phi(3S_{pkj;PC})-1\right)+1\right]\Big/2\right\} \tag{10}$$

where $S_{pkj;PC_i}$ represents the univariate measure of process yield index for the ith principal component. By analogy to equation (9), the overall process yield can be established as Yield$=2\Phi(3TS_{pk;PC}) - 1$.

Lee *et al.* (2002) inferred the asymptotic distribution for an estimate $\hat{S}_{pk}$ of the process yield index S_{pk}. An approximate $100(1-\alpha)\%$ confidence interval for S_{pk} is expressed as

$$\left(\hat{S}_{pk} - \frac{\sqrt{\hat{a}^2+\hat{b}^2}}{6\sqrt{n}\phi(3\hat{S}_{pk})}Z_{\alpha/2},\ \hat{S}_{pk} + \frac{\sqrt{\hat{a}^2+\hat{b}^2}}{6\sqrt{n}\phi(3\hat{S}_{pk})}Z_{\alpha/2}\right) \tag{11}$$

where

$$\begin{aligned}
\hat{S}_{pk} &= \frac{1}{3}\Phi^{-1}\left\{\frac{1}{2}\Phi\left(\frac{USL-\overline{X}}{S}\right)+\frac{1}{2}\Phi\left(\frac{\overline{X}-LSL}{S}\right)\right\} \\
&= \frac{1}{3}\Phi^{-1}\left\{\frac{1}{2}\Phi\left(\frac{1-\hat{C}_{dr}}{\hat{C}_{dp}}\right)+\frac{1}{2}\Phi\left(\frac{1+\hat{C}_{dr}}{\hat{C}_{dp}}\right)\right\}, \\
\hat{a} &= d\Big/\sqrt{2}S\left\{(1-\hat{C}_{dr})\phi\left(\frac{1-\hat{C}_{dr}}{\hat{C}_{dp}}\right)+(1+\hat{C}_{dr})\phi\left(\frac{1+\hat{C}_{dr}}{\hat{C}_{dp}}\right)\right\}, \\
\hat{b} &= \phi\left(\frac{1-\hat{C}_{dr}}{\hat{C}_{dp}}\right)-\phi\left(\frac{1+\hat{C}_{dr}}{\hat{C}_{dp}}\right),
\end{aligned}$$

where $Z_{\alpha/2}$ is the upper $100(\alpha/2)\%$ point of the standard normal distribution, and ϕ is the probability density function of the standard normal distribution. Applying the above formula, we can obtain an approximate $100(1-\alpha)\%$ lower confidence bound for S_{pk}, then an approximate $100(1-\alpha)\%$ lower confidence bound for the process yield can be obtained.

Equation (11) can be used to establish approximate lower confidence bound for the process yield in the case of single characteristic. However, how to estimate the process yield is more difficult for processes with multiple characteristics. By using the PCA method, such difficulty can be overcome. Applying the PCA method to equations (10) and (11), an approximate $100(1-\alpha)\%$ confidence interval for the combined index, $TS_{pk,PC}$, is expressed as

$$\frac{1}{3}\Phi^{-1}\left\{\left[\prod_{j=1}^{v}(2\Phi(3k_j)-1)+1\right]\Big/2\right\} \le TS_{pk,PC} \le \frac{1}{3}\Phi^{-1}\left\{\left[\prod_{1}^{v}(2\Phi(3\ell_j)-1)+1\right]\Big/2\right\}, \tag{12}$$

where

$$k_j = \hat{S}_{pkj;PC} - \frac{\sqrt{\hat{a}_{j;PC}^2+\hat{b}_{j;PC}^2}}{6\sqrt{n}\phi(3\hat{S}_{pkj;PC})}Z_{\alpha/2}m, \quad \ell_j = \hat{S}_{pkj;PC} + \frac{\sqrt{\hat{a}_{j;PC}^2+\hat{b}_{j;PC}^2}}{6\sqrt{n}\phi(3\hat{S}_{pkj;PC})}Z_{\alpha/2}.$$

Thus, an approximate $100(1-\alpha)\%$ lower confidence bound for $TS_{pk,PC}$ can be obtained, and an approximate $100(1-\alpha)\%$ lower confidence bound for the true process yield can also be obtained by using the one-to-one correspondence between $TS_{pk,PC}$ and the process yield (Yield $= 2\Phi(3TS_{pk;PC})-1$).

4. Application examples

Three examples are given below to illustrate the proposed methodology. We show how to calculate the lower confidence bound for process yield. In the first example, the case involves two quality characteristics. For the other two examples, the cases involve more than three quality characteristics.

4.1 *Example 1*

Chen (1994) discussed a bivariate normal example and employed Sultan (1986) bivariate process data ($n = 25$). Of particular interest were the brinell hardness (H) and the tensile strength (S) of a process. The specification limits for H and S were set at [112.7, 241.3] and [32.7, 73.3], respectively. The centre of the specifications was $T^T = [177, 53]$. The sample mean vector and sample covariance matrix were

$$\overline{X}^T = [177.2, 52.32] \quad \text{and} \quad S = \begin{bmatrix} 338 & 88.75 \\ 88.75 & 33.47414 \end{bmatrix}.$$

The process points and tolerance region is illustrated in figure 1.

By performing the principal components analysis, the eigenvecters and eigenvalues can be obtained. Table 2 shows the loading and eigenvalue of PCs using the principal component analysis. Testing the hypothesis H_0: $\lambda_1 = \lambda_2$ yields a value of $\chi_2^2 = 55.35$, which is quite significant at the 95% confidence level. Thus, we only used the first PC to evaluate the capability at 97% total variability. The principal components are $USL_{PC1} = 252.0660$, $LSL_{PC1} = 117.3279$, $\overline{X}_{PC1} = 184.7172$, $T_{PC1} = 184.6970$, and $S_{PC1} = 19.0257$. Referring to equation (11), $\hat{S}_{pk1;PC}$ can be calculated as 1.1803. Applying equation (12), the approximate 95% lower confidence bound for the combined index, $TS_{pk,PC}$, is 0.9058. Using the one-to-one

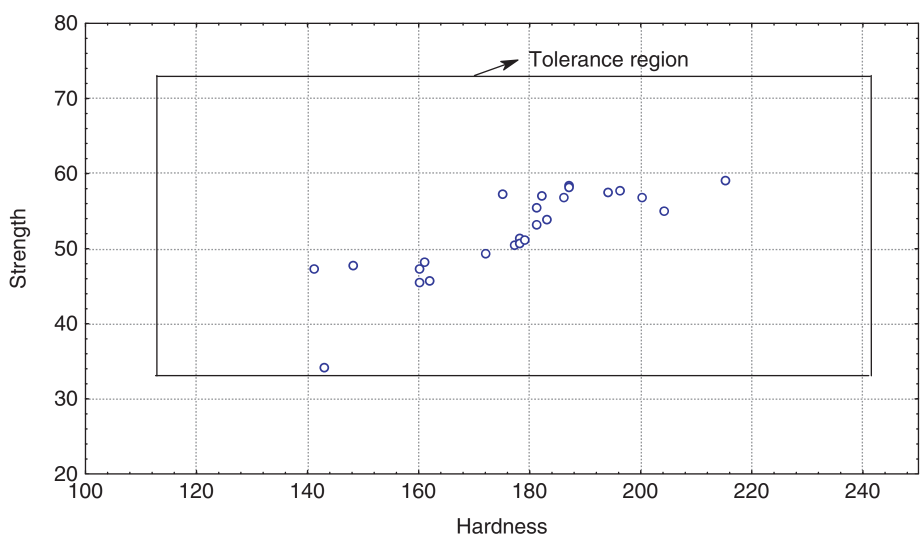

Figure 1. Process points and tolerance region for example 1.

Table 2. The results of PCA for example 1.

Variable	PC1 loading	PC2 loading
H	0.965389	−0.260814
S	0.260814	0.965389
Eigenvalue	361.9771	9.4970
% Explained of total variability	97.4434	2.5566

correspondence between $TS_{pk,PC}$ and the process yield (Yield $= 2\Phi(3TS_{pk;PC}) - 1$), the approximate 95% lower confidence bound for the true process yield is 0.993418. Notably, this process does not meet the process yield requirement.

4.2 *Example 2*

The previous study (Wang and Chen 1998) presented a trivariate quality control involving the joint control of the depth (D), the length (L) and the width (W) of a plastic product. Fifty observations are collected from a plastic production line. The specified limits for D, L, and W are set at [2.1, 2.3], [304.5, 305.1] and [304.5, 305.1], respectively. The specification of the target value is $T^T = [2.2, 304.8, 304.8]$. The p-value for Mardia's SW statistic is 0.32. Thus, the assumption of multivariate normality can not be rejected at 95% confidence level. The sample mean vector and sample covariance matrix were

$$\overline{X}^T = [2.1616, \quad 304.7182, \quad 3.4.7678] \quad \text{and}$$
$$S = \begin{bmatrix} 0.002051 & 0.000875 & 0.000656 \\ 0.000785 & 0.001717 & 0.001204 \\ 0.000656 & 0.001204 & 0.002034 \end{bmatrix}.$$

Figure 2 illustrate the process points and tolerance region.

By performing the principal components analysis, the eigenvecters and eigenvalues can be obtained. Table 3 shows the loading and eigenvalue of PCs using the principal component analysis. First, the test of the hypothesis H_0: $\lambda_1 = \lambda_2 = \lambda_3$, produced a value of $\chi^2_5 = 36.47$, which is significant at the 95% confidence level. That is, the hypothesis is rejected. Then, testing the hypothesis H_0: $\lambda_2 = \lambda_3$ produced a value of $\chi^2_2 = 8.19$, which is also significant at the 95% confidence level. Thus, we used the first two *PCs* to evaluate the capability at 89.04% total variability. The principal components are $USL_{PC1} = -368.1421$, $LSL_{PC1} = -368.9698$, $\overline{X}_{PC1} = -368.4682$, $T_{PC1} = -368.5560$, $S_{PC1} = 0.0609$, $USL_{PC2} = 216.8123$, $LSL_{PC2} = 216.5499$, $\overline{X}_{PC2} = 216.6794$, $T_{PC2} = 216.6811$ and $S_{PC2} = 0.0382$. Referring to equation (11), $\hat{S}_{pk1;PC}$ and $\hat{S}_{pk2;PC2}$ can be calculated as 1.7331 and 1.1450, respectively. Applying equation (12), the approximate 95% lower confidence bound for the combined index, $TS_{pk,PC}$, is 0.9566. Using the one-to-one correspondence between $TS_{pk,PC}$ and the process yield (Yield $= 2\Phi(3TS_{pk;PC}) - 1$), the approximate 95% lower confidence bound for the true process yield is 0.995894. Notably, this process does not meet the process yield requirement.

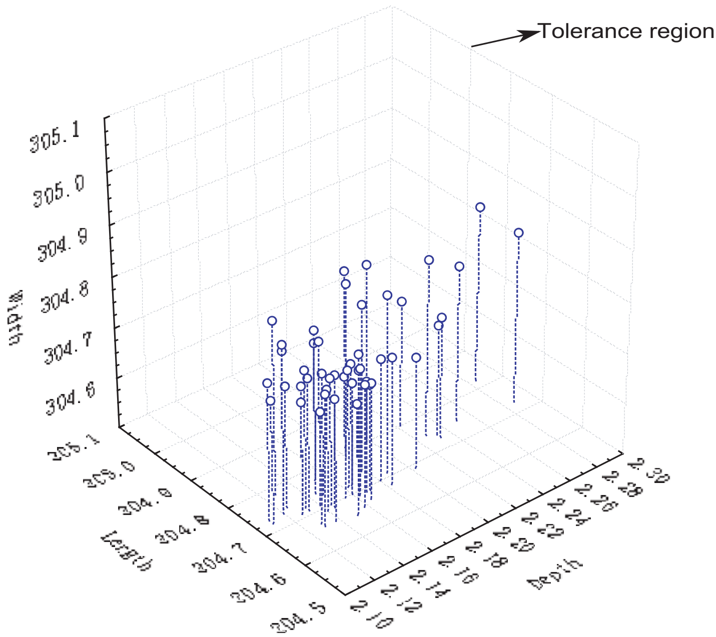

Figure 2. Process points and tolerance region for example 2.

Table 3. The results of the first two PCs for example 2.

Variable	PC1 loading	PC2 loading
D	−0.522185	−0.838481
L	−0.582407	0.217154
W	−0.622997	0.499794
Eigenvalue	0.003709	0.001457
% Explained of total variability	63.9253	25.1129

4.3 *Example 3*

Let we consider a real production process of multivariate normal distribution from an electronic thermos manufacturer located in Taiwan. One special type of thermos investigated has five target-the-best quality characteristics with unequal manufacturing specifications. Forty observations are generated from a multivariate normal distribution. The specification, target value, and the statistics of sample data are summarized as follows:

Variable	LSL	T	USL
X_1	5.598	6.220	6.842
X_2	606.5	680	753.5
X_3	0.215	0.354	0.493
X_4	31.5	35.0	38.5
X_5	30	40	50

Table 4. The results of the first two PCs for example 3.

Variable	PC1 loading	PC2 loading
1	0.015148	−0.000406
2	−0.985115	0.1684891
3	0.006121	0.005973
4	0.035063	0.026208
5	0.167490	0.985336
Eigenvalue	19.27390	13.50434
% Explained of total variability	57.8326	40.5207

$$\bar{X}^T = [6.4104, \quad 681.1656, \quad 0.3597, \quad 35.2364, \quad 41.0645],$$

$$S = \begin{bmatrix} 0.082811 & -0.286253 & 0.018106 & 0.043890 & 0.042138 \\ -0.286253 & 19.08823 & -0.101615 & -0.591470 & -0.93861 \\ 0.018106 & -0.101616 & 0.006509 & 0.030978 & 0.098379 \\ 0.043890 & -0.591470 & 0.030978 & 0.497112 & 0.446935 \\ 0.042138 & -0.93861 & 0.098379 & 0.446395 & 13.65237 \end{bmatrix}.$$

By performing the PCA, the eigenvectors and eigenvalues can be obtained. Table 4 shows the loading and eigenvalue of PCs using the principal component analysis. First, the test of the hypothesis H_0: $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5$, produced a value of χ^2_{14} =545.85, which is significant at the 95% confidence level. Second, the test of the hypothesis H_0: $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$, produced a value of χ^2_9 =242.72, which is significant at the 95% confidence level. Third, the test of the hypothesis H_0: $\lambda_1 = \lambda_2 = \lambda_3$, produced a value of $\chi^2_5 = 94.03$, which is significant at the 95% confidence level. That is, the hypothesis is rejected. Then, testing the hypothesis H_0: $\lambda_2 = \lambda_3$ produced a value of $\chi^2_2 = 79.60$, which is also significant at the 95% confidence level. Thus, we used the first two *PCs* to evaluate the capability at 98.35% total variability. The principal components are $USL_{PC1} = -591.2569$, $LSL_{PC1} = -732.4531$, $\overline{X}_{PC1} = -662.8138$, $T_{PC1} = -661.8550$, $S_{PC1} = 4.3902$, $USL_{PC2} = 177.2339$, $LSL_{PC2} = 132.5744$, $\overline{X}_{PC2} = 156.1566$, $T_{PC2} = 154.9042$ and $S_{PC2} = 3.6748$. Referring to equation (11), $\hat{S}_{pk1;PC}$ and $\hat{S}_{pk2;PC}$ can be calculated as 1.7331 and 1.7331, respectively. Applying equation (12), the approximate 95% lower confidence bound for the combined index, $TS_{pk,PC}$, is 1.6844. Using the one-to-one correspondence between $TS_{pk,PC}$ and the process yield (Yield $= 2\Phi(3TS_{pk;PC}) - 1$), the approximate 95% lower confidence bound for the true process yield is 0.999999. Notably, this process meets the process yield requirement.

5. Conclusions

Process yield is the most common and standard criteria for evaluating the quality of products manufactured. Process yield measure for processes with a single characteristic has been investigated extensively. However, process yield measure for processes with multiple quality characteristics is comparatively neglected. Assuring the process yield in processes with multiple characteristics to meet the requirement is important. So the proposition of a technique assuring the process yield is necessary in this field. In this paper, we proposed a generalized yield index, called $TS_{pk,PC}$, based on the yield index S_{pk} proposed by Boyles (1991), by using the PCA method. We also developed an approximate lower confidence bound (LCB) for the true process yield by using the S_{pk} through the principal component analysis (PCA). This methodology is easy to be understood and used. The proposed procedure can be used to determine whether their production meets the present yield requirement, and make a reliable decision.

Appendix

Theorem 1: *$\hat{S}^T_{pk}$ is defined as*

$$\hat{S}^T_{pk} = \frac{1}{3}\Phi^{-1}\left\{\left[\prod_{j=1}^{v}\left(2\Phi(3\hat{S}_{pkj}) - 1\right) + 1\right]\Big/ 2\right\},$$

where $\hat{S}_{pkj}$ denotes the estimator of S_{pkj}, $\hat{S}_{pkj} \sim N(S_{pkj}, (a_j^2 + b_j^2)/36n(\phi(3S_{pkj}))^2)$, and all $\hat{S}'_{pkj}s$ are mutually independent, then $\hat{S}^T_{pk}$ has the asymptotic normal distribution with the mean S^T_{pk} and variance

$$\frac{1}{36n(\phi(3S^T_{pk}))^2}\left(\sum_{j=1}^{v}\left\{(a_j^2 + b_j^2)\left[\frac{\prod_{i=1}^{v}\left(2\Phi(3S_{pki}) - 1\right)^2}{\left(2\Phi(3S_{pkj}) - 1\right)^2}\right]\right\}\right).$$

That is,

$$\hat{S}^T_{pk} \sim N\left(S^T_{pk}, \frac{1}{36n(\phi(3S^T_{pk}))^2}\left(\sum_{j=1}^{v}\left\{(a_j^2 + b_j^2)\left[\frac{\prod_{i=1}^{v}\left(2\Phi(3S_{pki}) - 1\right)^2}{\left(2\Phi(3S_{pkj}) - 1\right)^2}\right]\right\}\right)\right).$$

Proof: Applying the first-order expansion of v-variate Taylor,

$$\Rightarrow \quad f(X) = f(X_0) + \sum_{j=1}^{v}\frac{\partial f(X_0)}{\partial x_i}(x_i - x_{i0}),$$

where $X = (x_1, x_2, \ldots\ldots, x_v)$.

We take $v = 2$ for example to derive the asymptotic distribution of $\hat{S}^T_{pk}$.

Here

$$E(\hat{S}_{pkj}) = S_{pkj}, Var(\hat{S}_{pkj}) = \frac{a_j^2 + g_j^2}{36n(\phi(3S_pkj))^2}, \quad \forall j = 1, 2.$$

From the definition, we have

$$\hat{S}^T_{pk} = f(S_{pk1}, S_{pk2}) + \frac{\partial f(S_{pk1}, S_{pk2})}{\partial \hat{S}_{pk1}}(\hat{S}_{pk1} - S_{pk1}) + \frac{\partial f(S_{pk1}, S_{pk2})}{\partial \hat{S}_{pk2}}(\hat{S}_{pk2} - S_{pk2}).$$

Then, we have

$$\begin{aligned}
E(\hat{S}_{pk}^T) &= E(f(S_{pk1}, S_{pk2})) + E\left(\frac{\partial f(S_{pk1}, S_{pk2})}{\partial \hat{S}_{pk1}}(\hat{S}_{pk1} - S_{pk1})\right) \\
&\quad + E\left(\frac{\partial f(S_{pk1}, S_{pk2})}{\partial \hat{S}_{pk2}}(\hat{S}_{pk2} - S_{pk2})\right) \\
&= f(S_{pk1}, S_{pk2}) = S_{pk}^T = \frac{1}{3}\Phi^{-1}\left\{\left[(2\Phi(3S_{pk1}) - 1)(2\Phi(3S_{pk2}) - 1) + 1\right]/2\right\}.
\end{aligned}$$

$$Var(\hat{S}_{pk}^T) = \left(\frac{\partial f(S_{pk1}, S_{pk2})}{\partial \hat{S}_{pk1}}\right)^2 Var(\hat{S}_{pk1}) + \left(\frac{\partial f(S_{pk1}, S_{pk2})}{\partial \hat{S}_{pk2}}\right)^2 Var(\hat{S}_{pk2}).$$

$$\because f(\hat{S}_{pk1}, \hat{S}_{pk2}) = \frac{1}{3}\Phi^{-1}\left\{\left[(2\Phi(3\hat{S}_{pk1}) - 1)(2\Phi(3\hat{S}_{pk2}) - 1) + 1\right]/2\right\}.$$

We have

$$\frac{\partial f(\hat{S}_{pk1}, \hat{S}_{pk2})}{\partial \hat{S}_{pk1}} = \frac{(2\Phi(3\hat{S}_{pk2}) - 1)\phi(3\hat{S}_{pk1})}{\phi\left\{\Phi^{-1}\left\{\left[(2\Phi(3\hat{S}_{pk1}) - 1)(2\Phi(3\hat{S}_{pk2}) - 1) + 1\right]/2\right\}\right\}}$$

$$\therefore \frac{\partial f(S_{pk1}, S_{pk2})}{\partial \hat{S}_{pk1}} = \frac{(2\Phi(3S_{pk2}) - 1)\phi(3S_{pk1})}{\phi\left\{\Phi^{-1}\left\{\left[(2\Phi(3S_{pk1}) - 1)(2\Phi(3S_{pk2}) - 1) + 1\right]/2\right\}\right\}}.$$

Similarly, we have

$$\frac{\partial f(S_{pk1}, S_{pk2})}{\partial \hat{S}_{pk2}} = \frac{(2\Phi(3S_{pk1}) - 1)\phi(3S_{pk2})}{\phi\left\{\Phi^{-1}\left\{\left[(2\Phi(3S_{pk1}) - 1)(2\Phi(3S_{pk2}) - 1) + 1\right]/2\right\}\right\}}.$$

So,

$$Var(\hat{S}_{pk}^T) = \frac{1}{36n(\phi(3S_{pk}^T))^2}\left[(a_1^2 + b_1^2)(2\Phi(3S_{pk2}) - 1)^2 + (a_2^2 + b_2^2)(2\Phi(3S_{pk1}) - 1)^2\right].$$

By Central Limit Theorem $=> \hat{S}_{pk}^T$ has the asymptotic normal distribution with the mean S_{pk}^T and variance

$$\frac{1}{36n(\phi(3S_{pk}^T))^2}\left[(a_1^2 + b_1^2)(2\Phi(3S_{pk2}) - 1)^2 + (a_2^2 + b_2^2)(2\Phi(3S_{pk1}) - 1)^2\right].$$

Similarly, consider v variables, the asymptotic distribution of $\hat{S}_{pk}^T$ can be derived as

$$\hat{S}_{pk}^T \sim N\left(S_{pk}^T, \frac{1}{36n(\phi(3S_{pk}^T))^2}\left(\sum_{j=1}^{v}\left\{(a_j^2 + b_j^2)\left[\frac{\prod_{i=1}^{v}\left(2\Phi(3S_{pki}) - 1\right)^2}{\left(2\Phi(3S_{pkj}) - 1\right)^2}\right]\right\}\right)\right).$$

Theorem 2: *Let Σ be the covariance matrix associated with the random vector*

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_v \end{bmatrix}.$$

Let Σ have the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_v$ & eigenvectors $e_1, \ldots, e_v$. Then the ith principal component variable is given by $y_i = e_i' X = e_{i1}x_1 + \cdots + e_{iv}x_v$, $i = 1, 2, \ldots, v$.

With these choices, $\mathrm{var}(y_i) = e_i' \Sigma e_i = \lambda_i$, $i = 1, 2, \ldots, v$ *and* $\mathrm{cov}(y_i, y_k) = e_i' \Sigma e_k = 0$, $i \neq k$.

Proof: The proof can be found in Johnson and Wichern (2002) on page 428.

Acknowledgements

The authors wish to gratefully acknowledge the referees of this paper who helped to clarify and improve the presentation.

References

Anderson, T.W., Asymptotic theory for principal component analysis. *Ann. Math. Stat.*, 1963, **34**, 122–148.
Boyles, R.A., The Taguchi capability index. *J. Qual. Technol.*, 1991, **23**, 17–26.
Chan, L.K., Cheng, S.W. and Spiring, F.A., A multivariate measure of process capability. *J. Model. Simul.*, 1991, **11**, 1–6.
Chen, H., A multivariate process capability index over a rectangular solid tolerance zone. *Stat. Sin.*, 1994, **4**, 749–758.
Chen, K.S., Pearn, W.L. and Lin, P.C., Capability measures for processes with multiple characteristics. *Qual. Reliab. Engng Int.*, 2003, **19**, 101–110.
Jackson, J.E., Principal component and factor analysis: part I—principal components. *J. Qual. Technol.*, 1980, **12**, 201–213.
Johnson, R.A. and Wichern, D.E., edition. in *Applied Multivariate Statistical Analysis*, fifth edn, pp. 428, 2002 (Prentice Hall: New Jersey).
Kotz, S. and Johnson, N.L., Process capability indices—a review, 1992–2000. *J. Qual. Technol.*, 2002, **34**, 2–19.
Lee, J.C., Hung, H.N., Pearn, W.L. and Kueng, T.L., On the distribution of the estimated process yield index S_{pk}. *Qual. Reliab. Engng Int.*, 2002, **18**, 111–116.
Pearn, W.L., Kotz, S. and Johnson, N.L., Distributional and inferential properties of process capability indices. *J. Qual. Technol.*, 1992, **24**, 216–231.
Shahriari, H., Hubele, N.F. and Lawrence, F.P, A multivariate process capability vector. in *Proceedings of the 4th Industrial Engineering Research Conference*. Institute of Industrial Engineers, Norcross, Georgia, USA, pp. 304–309/24–25 May, 1995.
Sultan, T.I., An acceptance chart for raw materials of two correlated properties. *Quality Assurance*, 1986, **12**, 70–72.
Taam, W., Subbaiah, P. and Liddy, J.W., A note on multivariate capability indices. *J. Appl. Stat.*, 1993, **20**, 339–351.
Wang, F.K., Capability index using principal component analysis. *Quality Engineering*, 1998, **11**, 21–27.
Wang, F.K. and Du, T.C.T., Using principal component analysis in process performance for multivariate data. *Omega, Int. J. Manag. Sci.*, 2000, **28**, 185–194.